Università degli studi di Trieste,
Department of Physics

Morphological Classification of Galaxies by Deep Learning Techniques

Department of Physics, Università degli studi di Trieste
Bachelor Degree in Physics
18th October 2019

Candidate: Sebastiano Zagatti

Supervisor: Prof. Mauro Messerotti
Co-Supervisor: Dott. Daniele Tavagnacco

Scheme of the presentation

1. Introduction
2. Morphological Classification of Galaxies
3. Machine Learning (ML) and Deep Learning (DL)
4. Image Analysis
5. Results
6. Conclusions and Perspectives

Università degli studi di Trieste,
Department of Physics

Introduction

The number of astronomical images, in particular galaxies, has been increasing in the last decade.

A tool for automatic morphological classification of galaxies is needed.

The goal of this thesis is to identify and test a method for automatic morphological classification of galaxies.

Deep Learning with Convolutional Neural Networks is a suitable method.

2. Morphological Classification of Galaxies

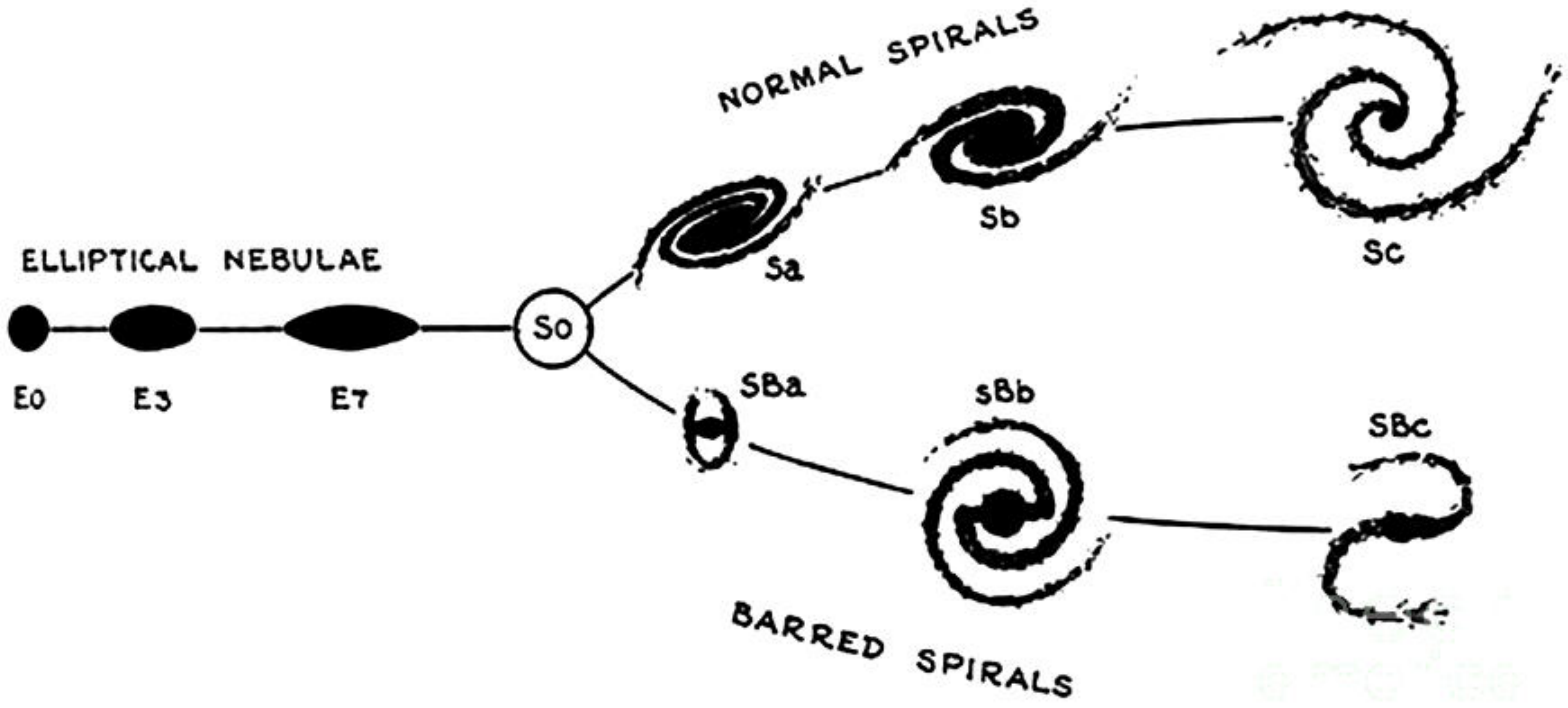
- The Hubble Sequence
- The de Vaucouleurs System

Università degli studi di Trieste,
Department of Physics

The Hubble Sequence

The first morphological classification of galaxies was made by Edwin Hubble in 1926.
Three main types can be identified:

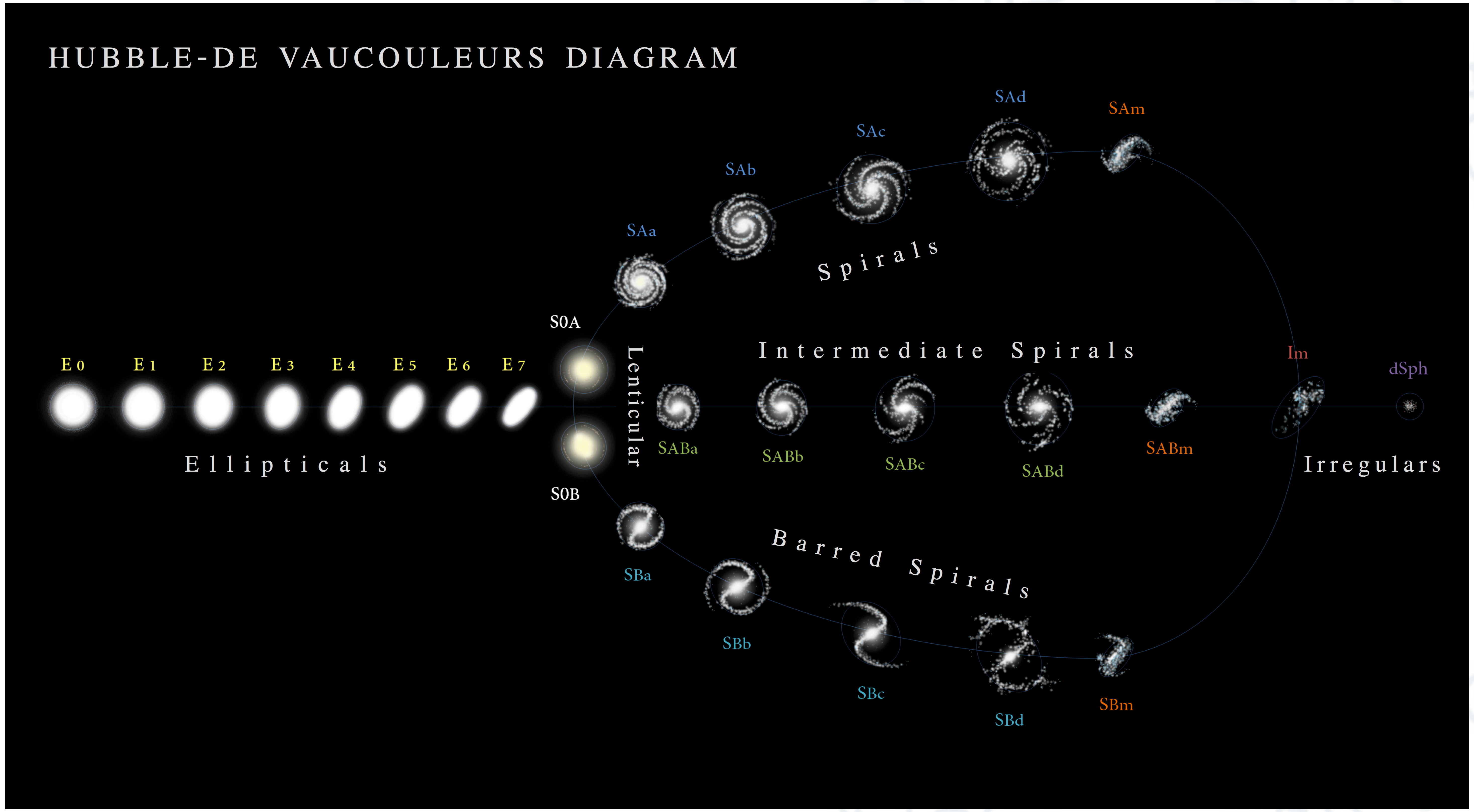
1. Elliptical Galaxies;
2. Lenticular Galaxies;
3. Spiral Galaxies.



De Vaucouleurs System

Extension of Hubble's Sequence by adding Sd, Sm and Im types,

Introduction of an intermediate sequence: "weakly barred" galaxies (SAB).



3. Machine Learning and Deep Learning

- Artificial Neural Networks
- Machine Learning vs Deep Learning
- The GPU Accelerated Computer
- The Implemented DL MATLAB® Script

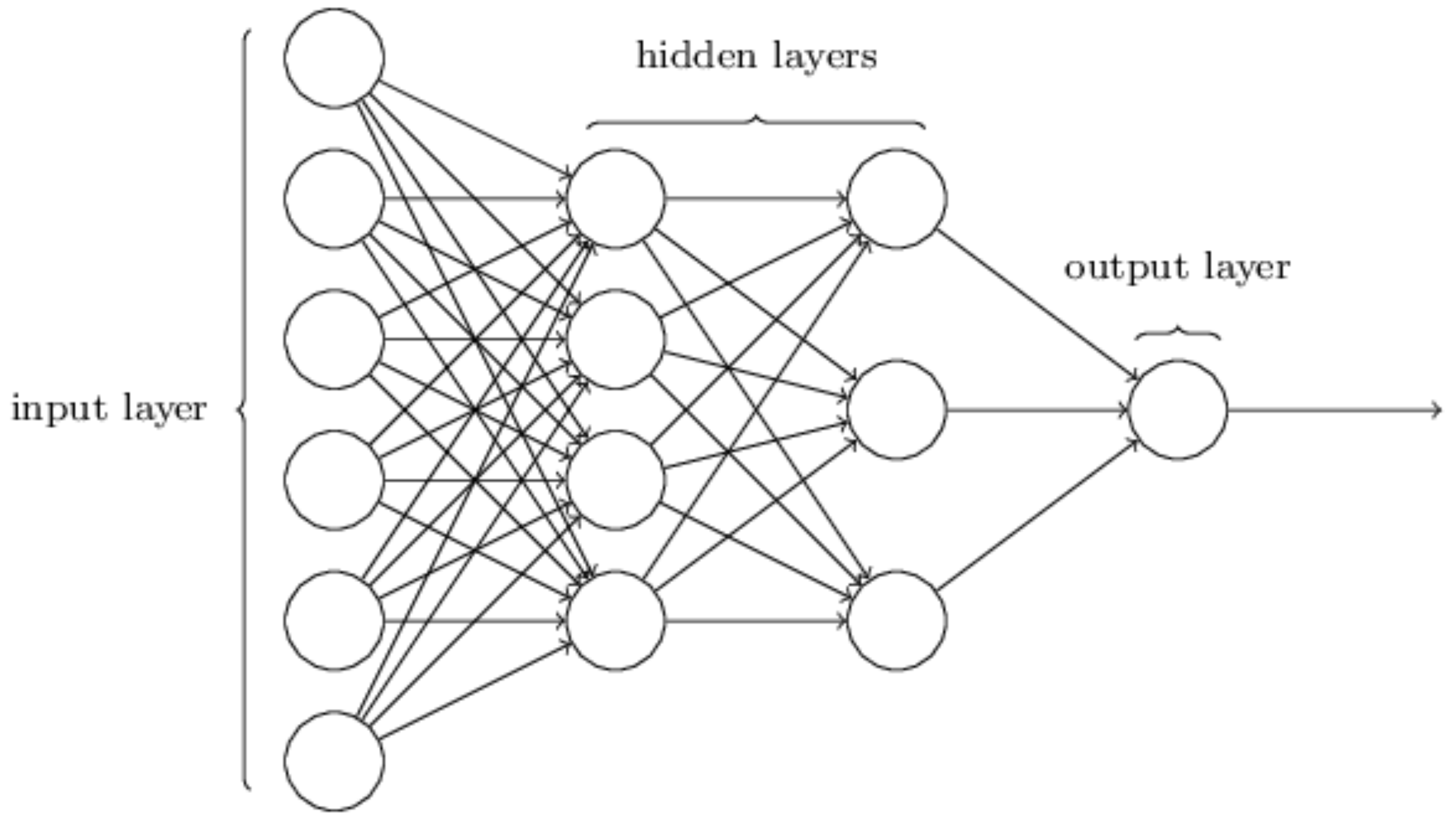
Università degli studi di Trieste,
Department of Physics

Artificial Neural Networks

Mathematical computing systems that learn to perform tasks by considering examples without being programmed with task-specific rules.

Series of interconnected nodes (artificial neurons) that receive an input, process it and produce an output signal, passing it through the next one.

Neurons are aggregated into layers, performing different transformation on the data.

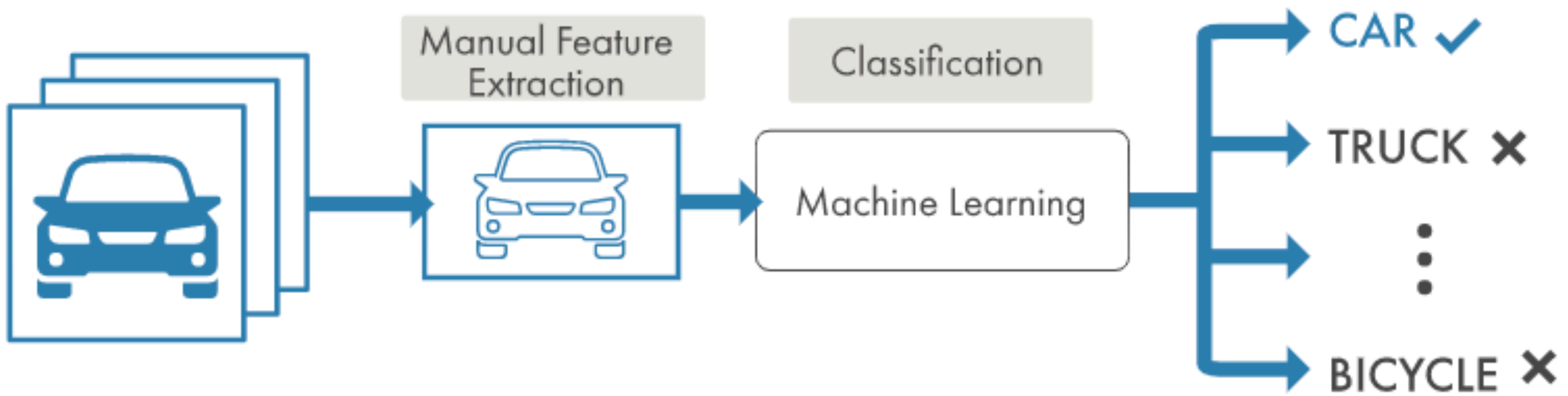


Università degli studi di Trieste,
Department of Physics

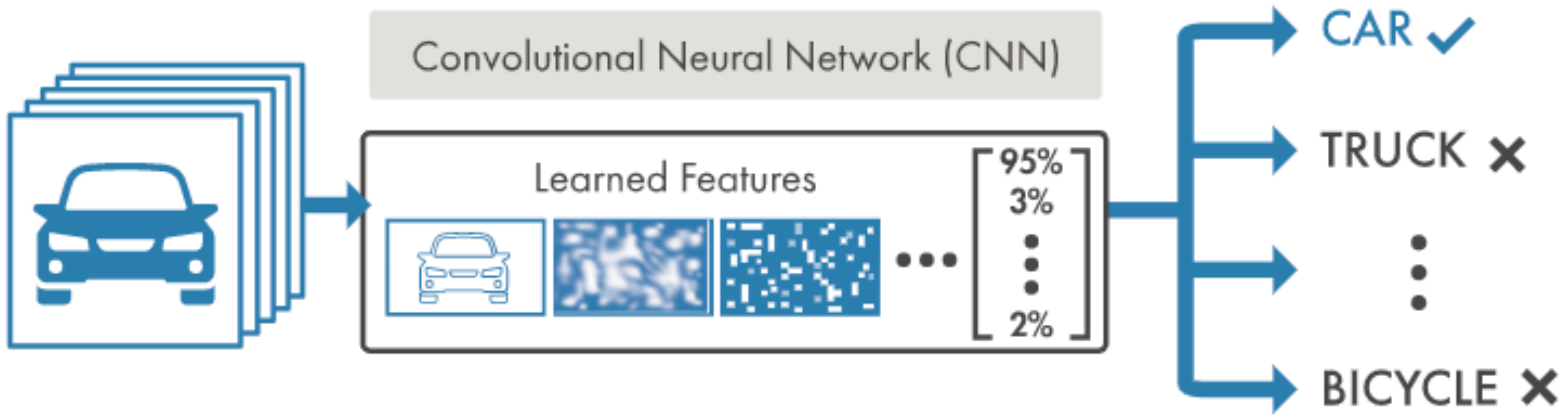
Machine Learning vs Deep Learning

In ML image classification, the relevant features of the images are selected manually. The model refers to those features when analyzing new inputs.

In DL image classification, the images are directly elaborated by the algorithm which automatically identifies the features and then refers to those when analyzing new inputs.



Machine Learning



Deep Learning

Università degli studi di Trieste,
Department of Physics

The GPU Accelerated Computer

Deep Learning involves high computing demand that can only be achieved by high-performance GPUs.

The HP Proliant SX40 GPU accelerated computer that was used has four Tesla V100 Nvidia GPUs with a capacity of 16 GB.

This server is part of the computing infrastructure of the COSMOS project (ASI) and running time was kindly granted by the INAF-Astronomical Observatory of Trieste.

Università degli studi di Trieste,
Department of Physics

The Implemented DL MATLAB® Script

Built on a MATLAB® template designed to classify the MNIST handwritten digits database (10000 [28px x 28px x 1] images).

Università degli studi di Trieste,
Department of Physics

The Implemented DL MATLAB® Script

The script loads the images and divides them in two sets:

- Training Set;
- Validation Set.

Università degli studi di Trieste,
Department of Physics

The Implemented DL MATLAB® Script

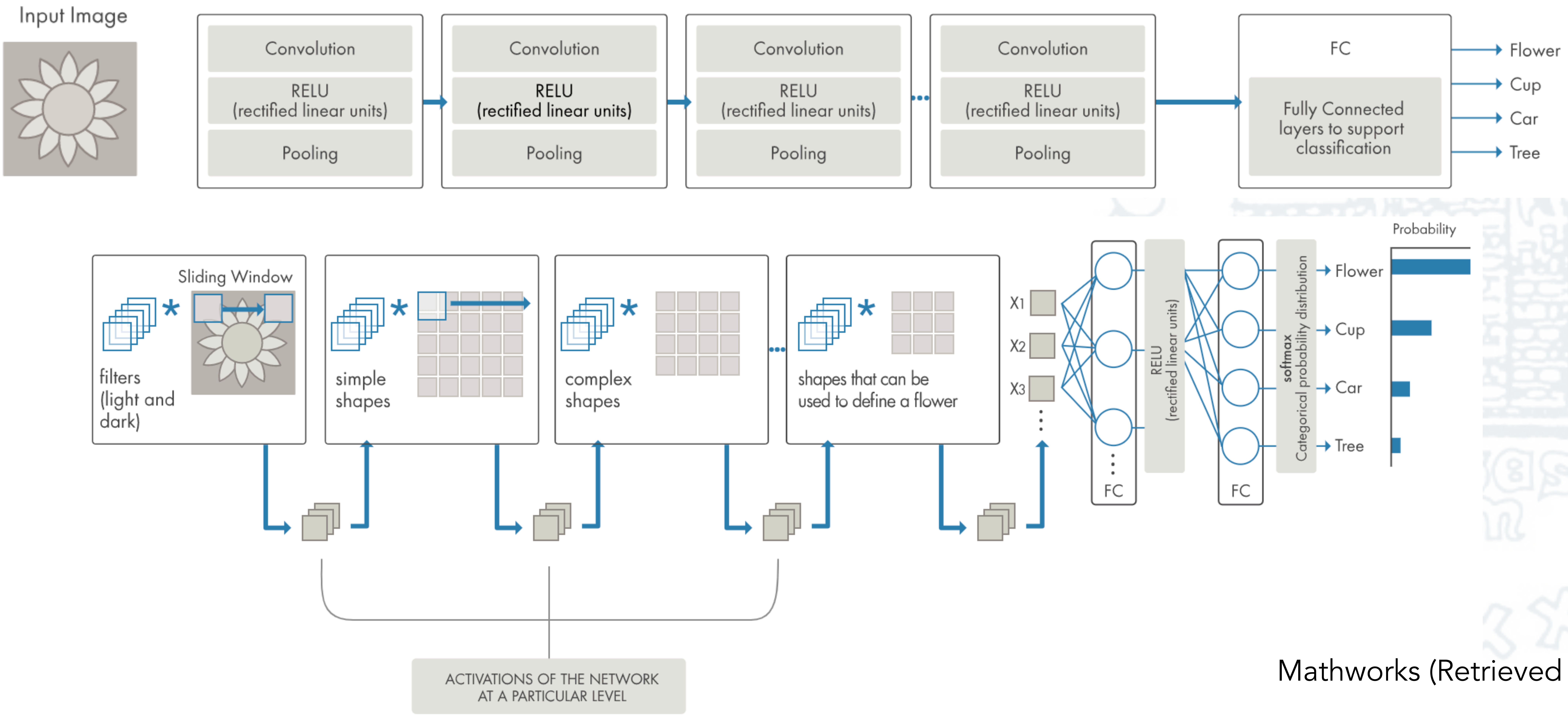
Convolutional Neural Networks (CNN)

One of the most popular network architectures for Deep Learning.

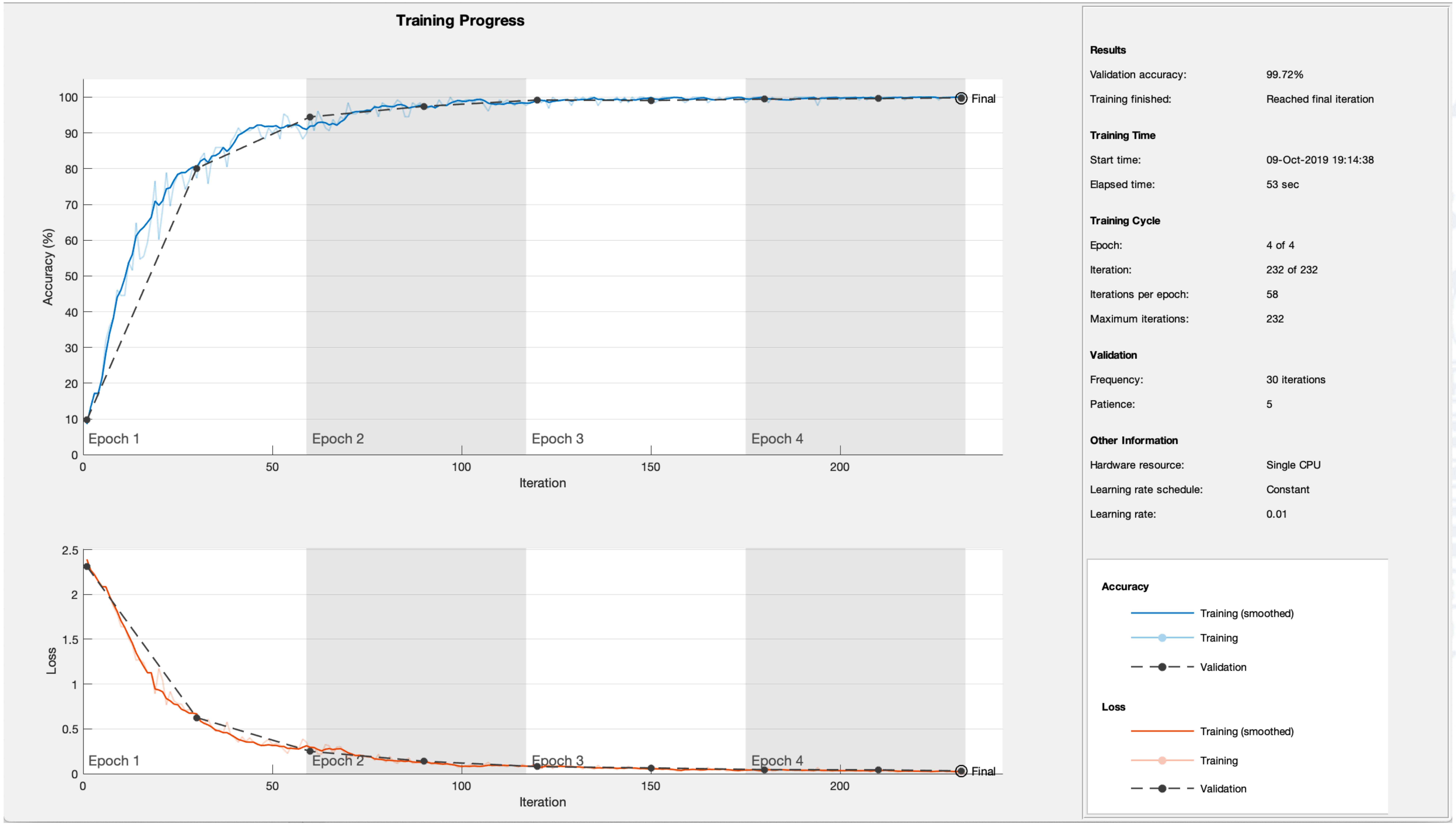
Made up of several 2D layers, providing a well suited framework to process images and carry out automatic object recognition and classification.

Università degli studi di Trieste,
Department of Physics

Sample scheme of a CNN used for object classification in images



Mathworks (Retrieved 2019)



Training Progress
Accuracy (%)
Iteration
Epoch 1
Epoch 2
Epoch 3
Epoch 4
Final
Loss
Results
Validation accuracy: 99.72%
Training finished: Reached final iteration
Training Time
Start time: 09-Oct-2019 19:14:38
Elapsed time: 53 sec
Training Cycle
Epoch: 4 of 4
Iteration: 232 of 232
Iterations per epoch: 58
Maximum iterations: 232
Validation
Frequency: 30 iterations
Patience: 5
Other Information
Hardware resource: Single CPU
Learning rate schedule: Constant
Learning rate: 0.01
Accuracy
Training (smoothed)
Training
Validation
Loss
Training (smoothed)
Training
Validation

Università degli studi di Trieste,
Department of Physics

The Implemented DL MATLAB® Script

Some changes to the script were made:

1. Addition of a Max-Pooling Layer;

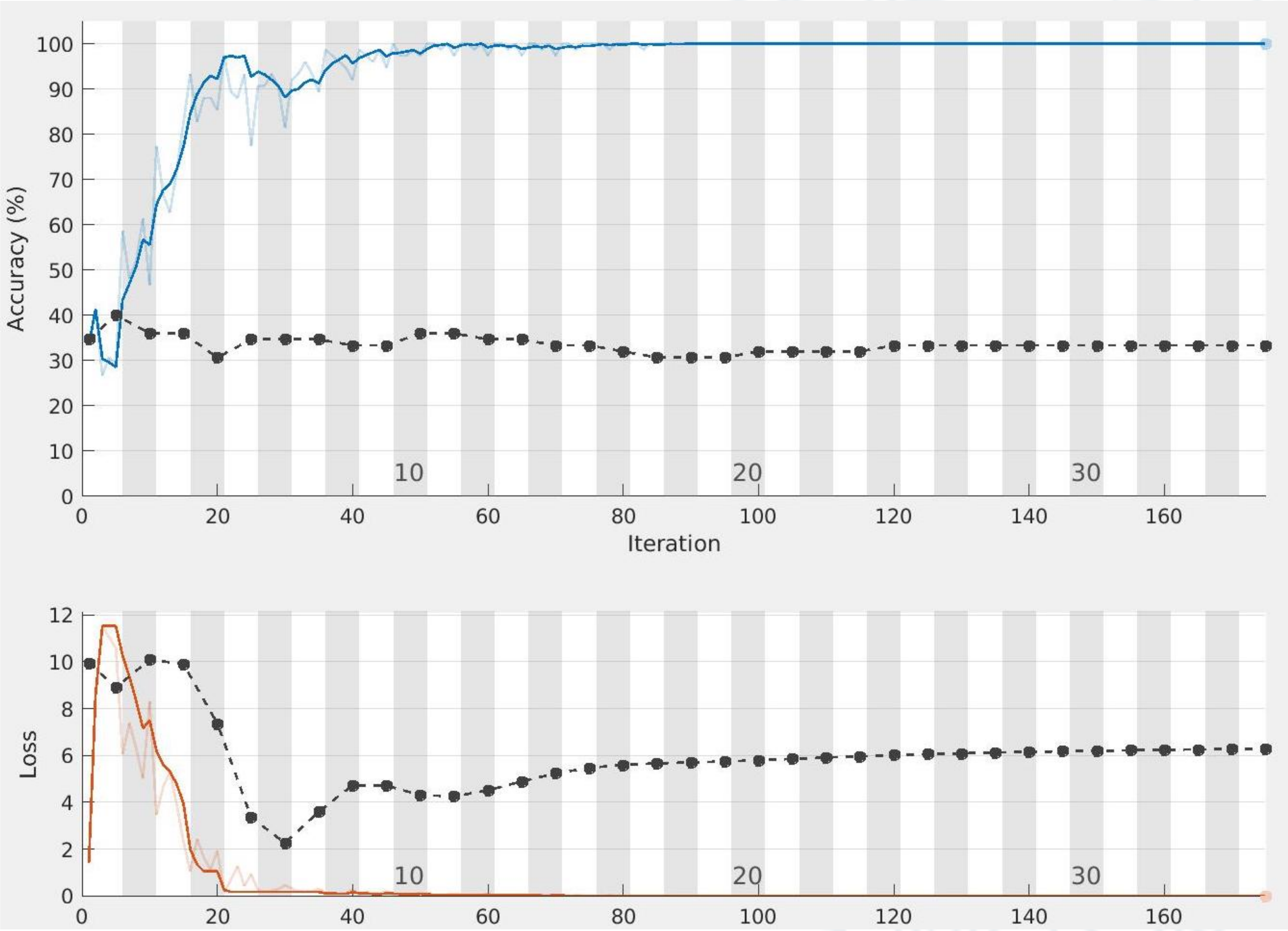
Università degli studi di Trieste,
Department of Physics

Results of the Modified CNN:

1. Input Layer
2. Convolutional Layer
3. Batch Normalization Layer
4. ReLU Layer
5. Max-Pooling Layer
6. Convolutional Layer
7. Batch Normalization Layer
8. ReLU Layer
9. Fully Connected Layer
10. Softmax Layer
11. Classification Layer

Convolutional Block (layers 2–4)

Convolutional Block (layers 6–8)



Three Classes - Accuracy: 33% External test: 36%

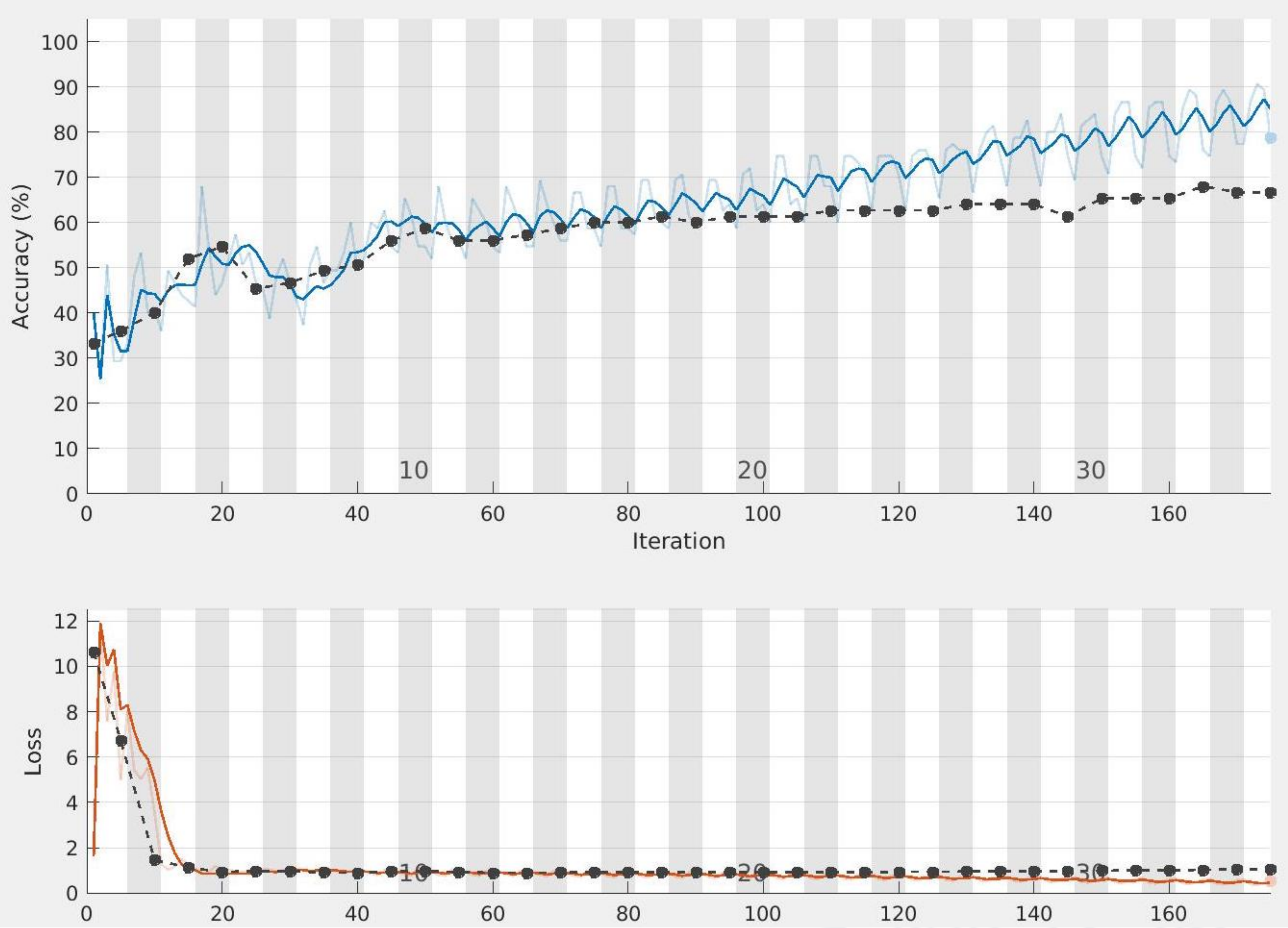
Università degli studi di Trieste,
Department of Physics

Results of the Modified CNN:

1. Input Layer
2. Convolutional Layer
3. Batch Normalization Layer
4. ReLU Layer
5. Max-Pooling Layer
6. Convolutional Layer
7. Batch Normalization Layer
8. ReLU Layer
9. Max-Pooling Layer
10. Fully Connected Layer
11. Softmax Layer
12. Classification Layer

Convolutional Block (layers 2–4)

Convolutional Block (layers 6–8)



Accuracy: 66.67% External test: 56%

Università degli studi di Trieste,
Department of Physics

The Implemented DL MATLAB® Script

2. Considering the differences between the MNIST and the used astronomical dataset the number of layers was reduced from 15 to 8;

1. Input Layer
2. Convolutional Layer
3. Batch Normalization Layer
4. ReLU Layer
5. Max-Pooling Layer
6. Fully Connected Layer
7. Softmax Layer
8. Classification Layer

Convolutional Block

Università degli studi di Trieste,
Department of Physics

The Implemented DL MATLAB® Script

3. Random selection of 125 training images and 25 validation images;

4. Additional validation process to test the already-trained network on 25 random images;

5. Selection of the Multi-GPU computational environment.

4. Image Analysis

- Image Dataset Selection
- Image Pre-Processing

Università degli studi di Trieste,
Department of Physics

Image Dataset Selection

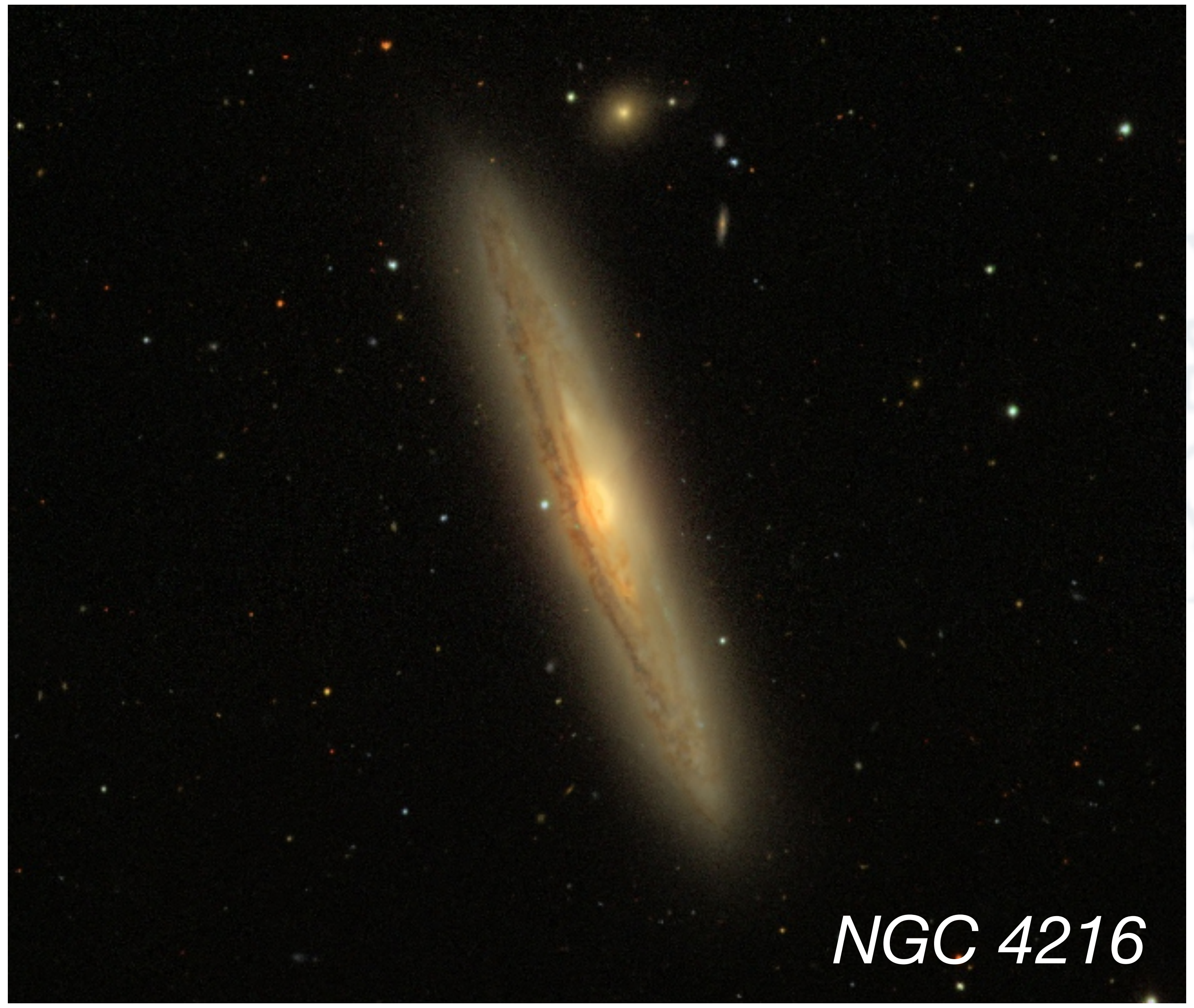
1. NASA/IPAC Extragalactic Database (NED);
2. Strasbourg Astronomical Data Center;
3. SIMBAD astronomical database;
4. Aladin interactive software sky atlas;
5. Sloan Digital Sky Survey (SDSS) up to DR9.

Messier 109

Image Dataset Selection

Created database [649 px x 550 px x 3]:

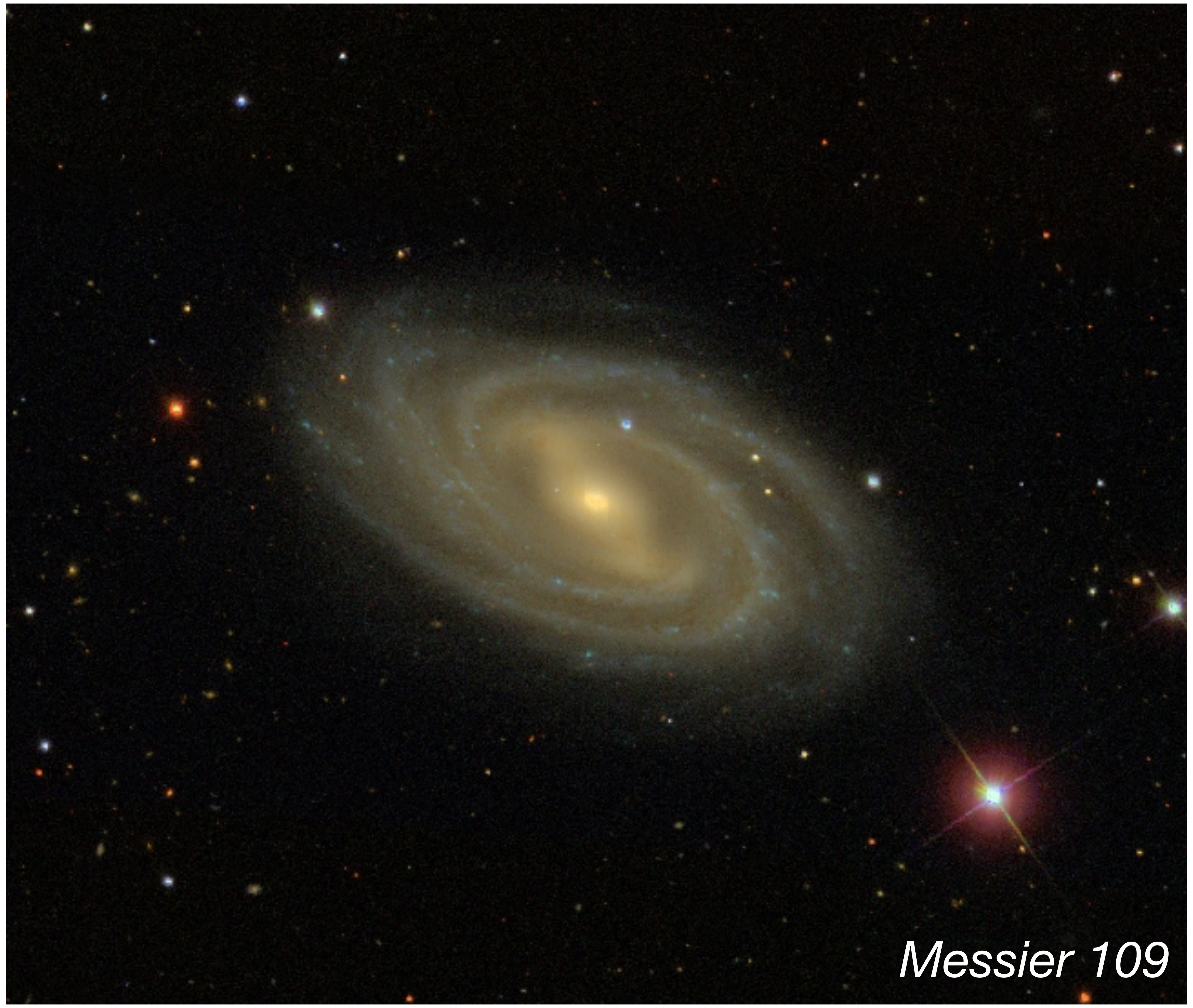
- 625 elliptical galaxies;
- 175 spiral galaxies;
- 322 barred spirals;
- 189 edge-on galaxies.



Università degli studi di Trieste,
Department of Physics

Image Pre-Processing

1. Grayscale Conversion

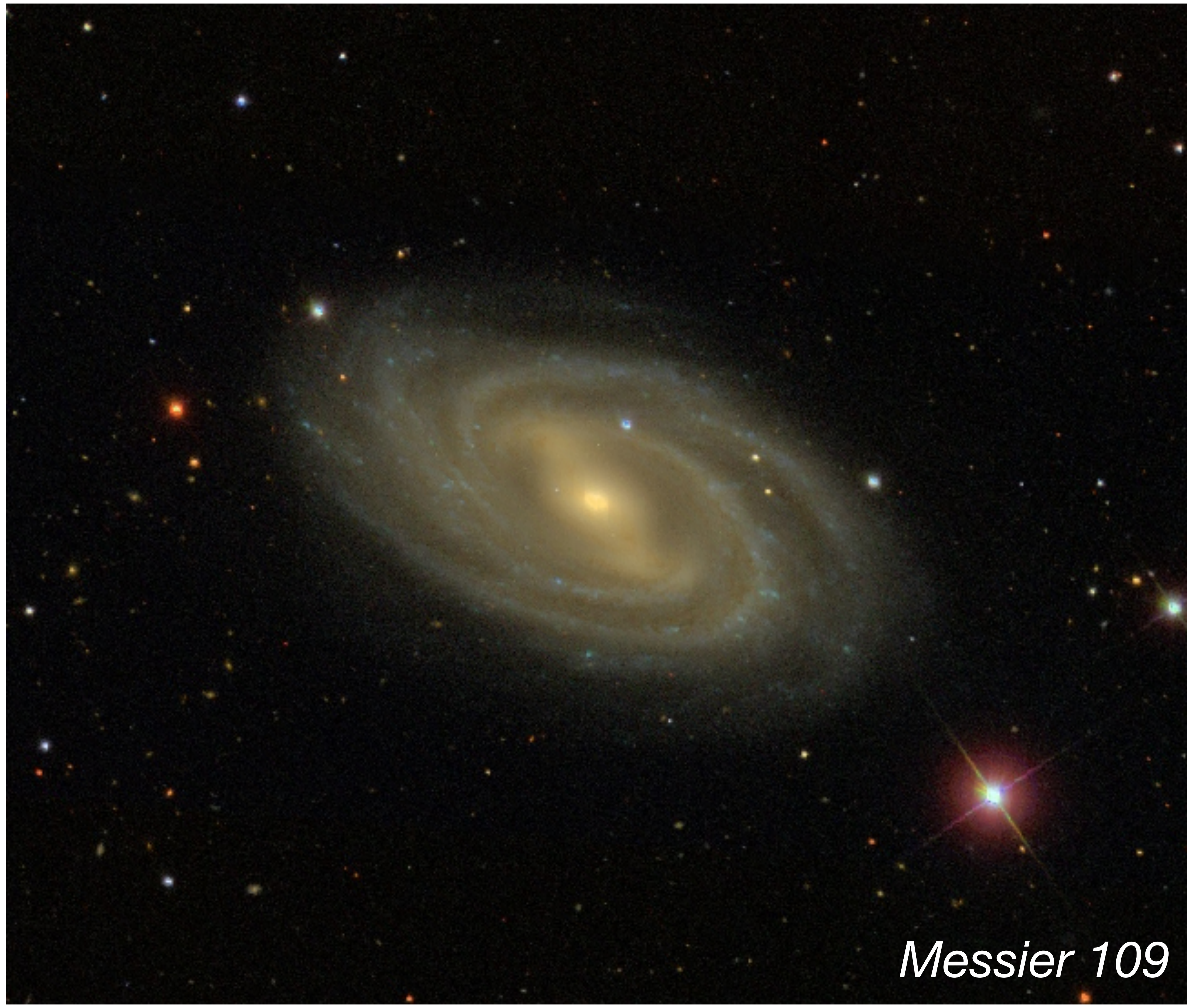


Messier 109

Università degli studi di Trieste,
Department of Physics

Image Pre-Processing

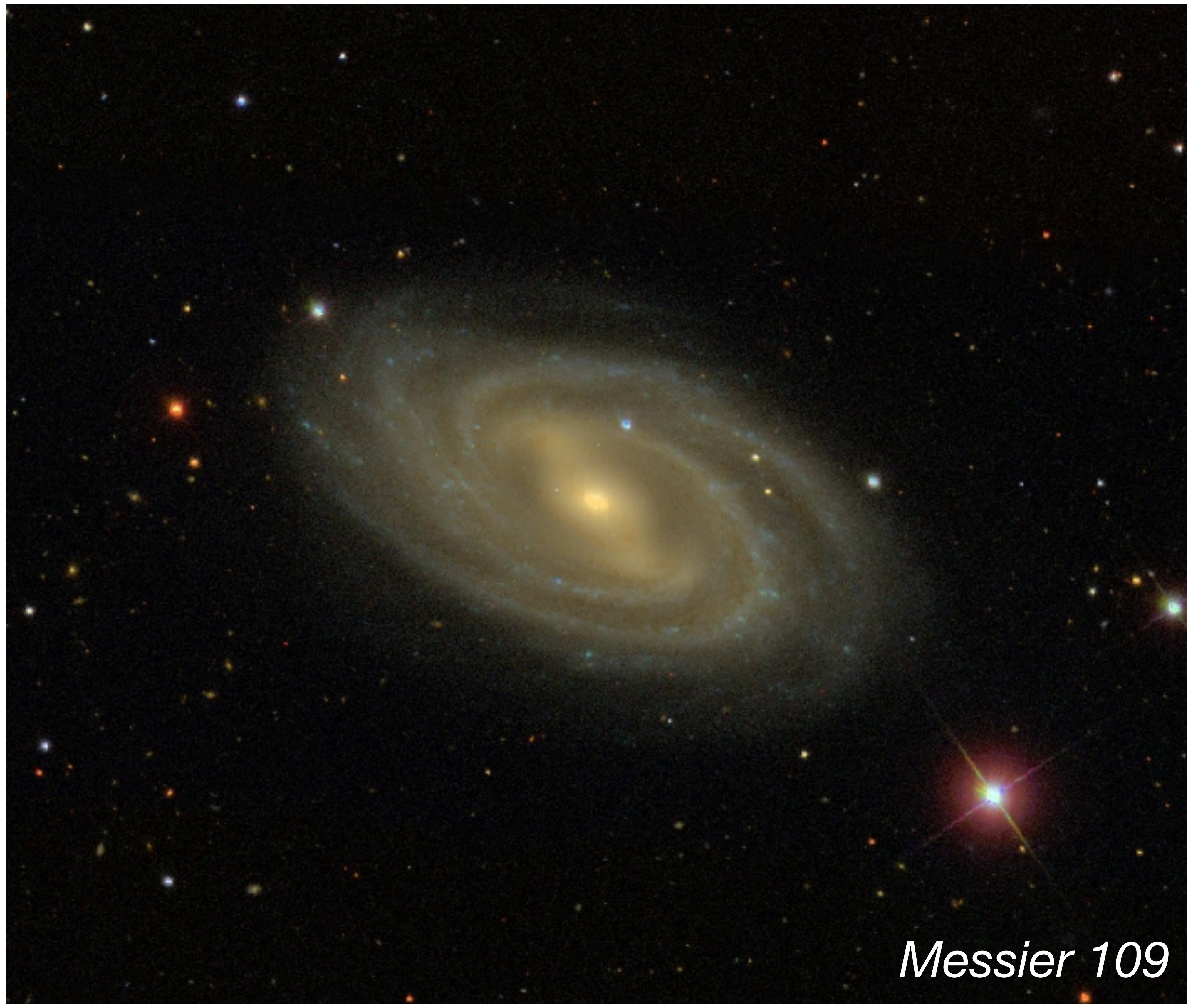
2. Sharpening High-Pass Gaussian Profile Filter

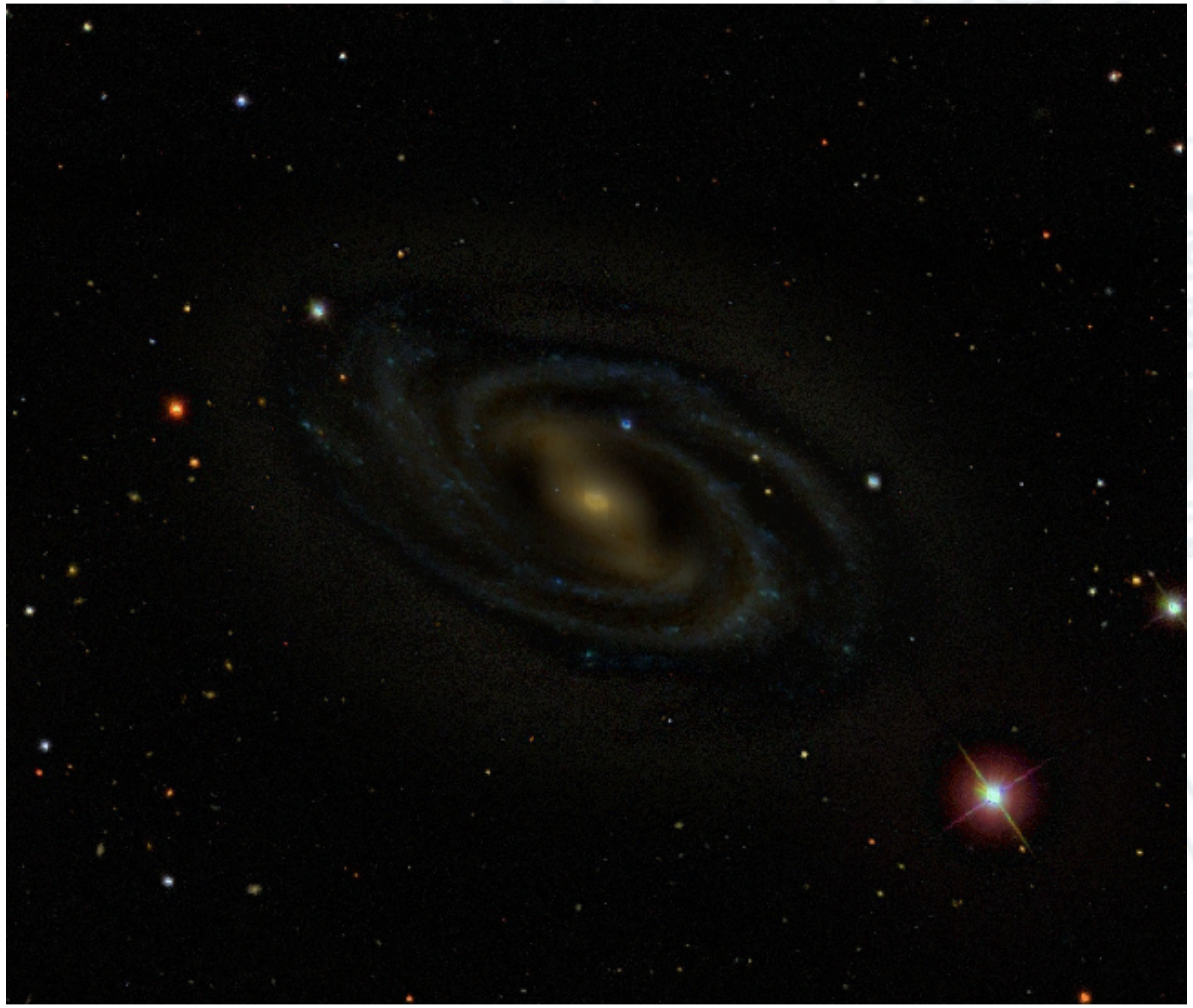


Università degli studi di Trieste,
Department of Physics

Image Pre-Processing

3. High-Pass FFT Gaussian Profile Filter

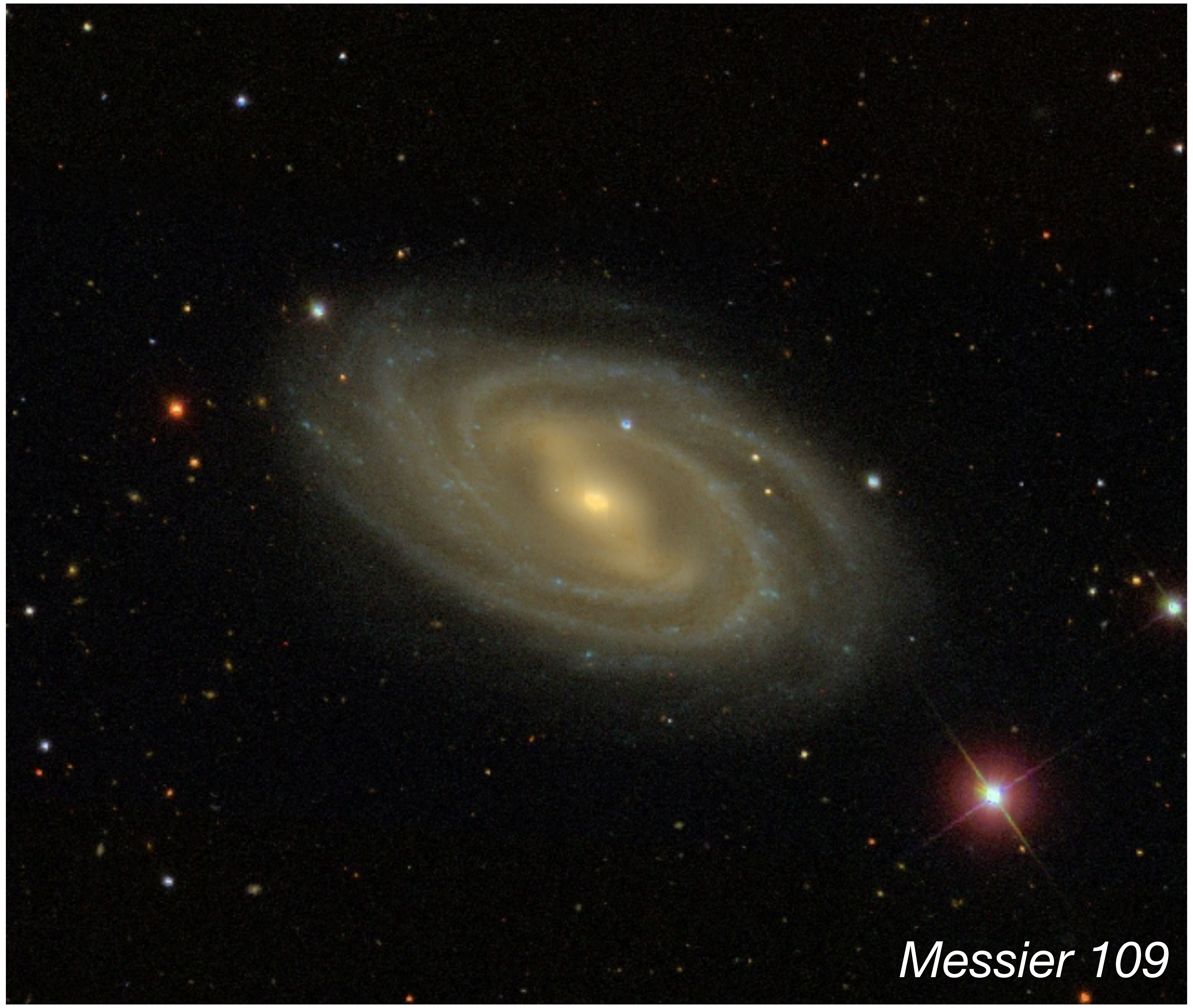


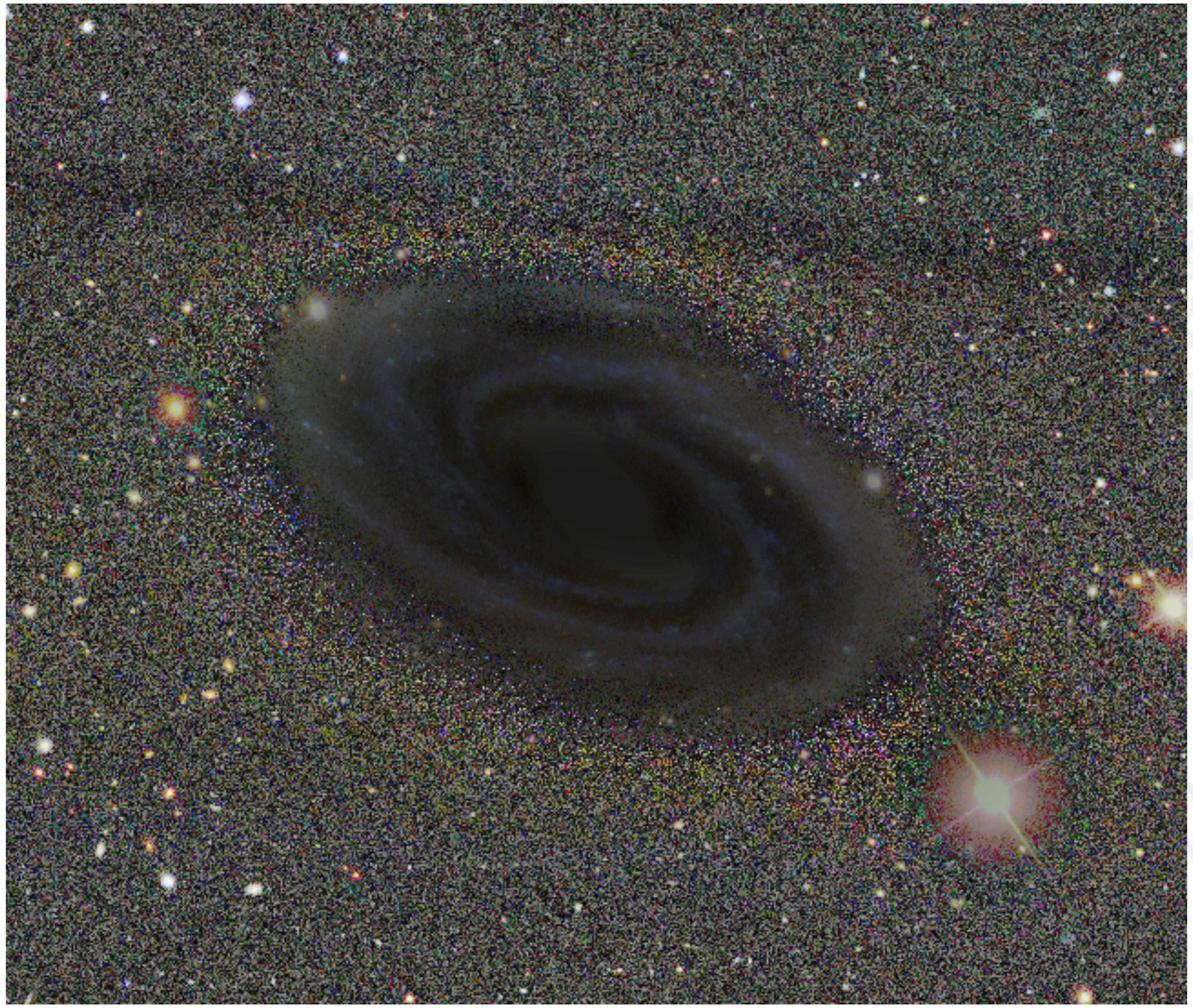


Università degli studi di Trieste,
Department of Physics

Image Pre-Processing

4. Histogram Equalization and High-Pass FFT Filter



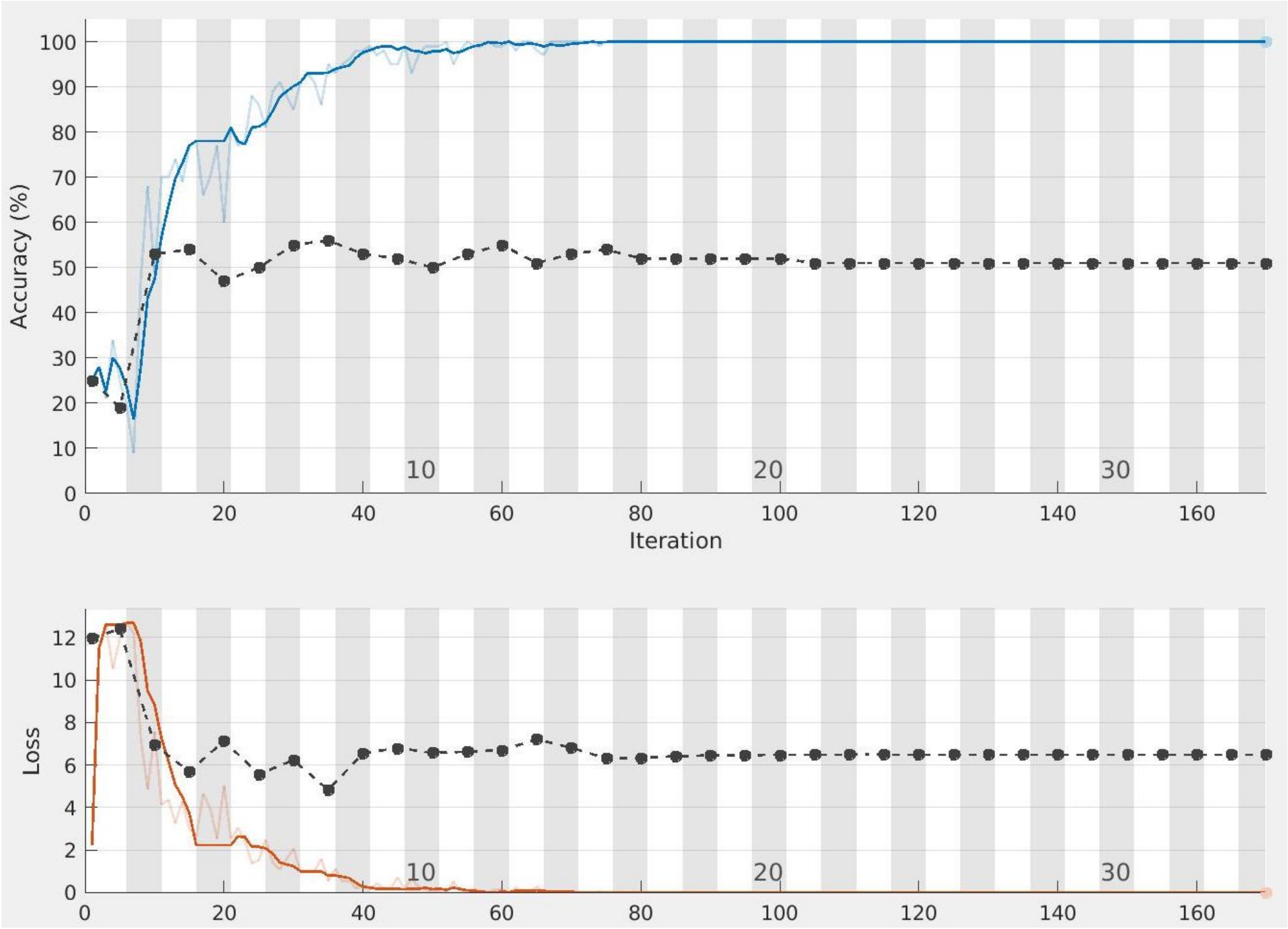


5. Results

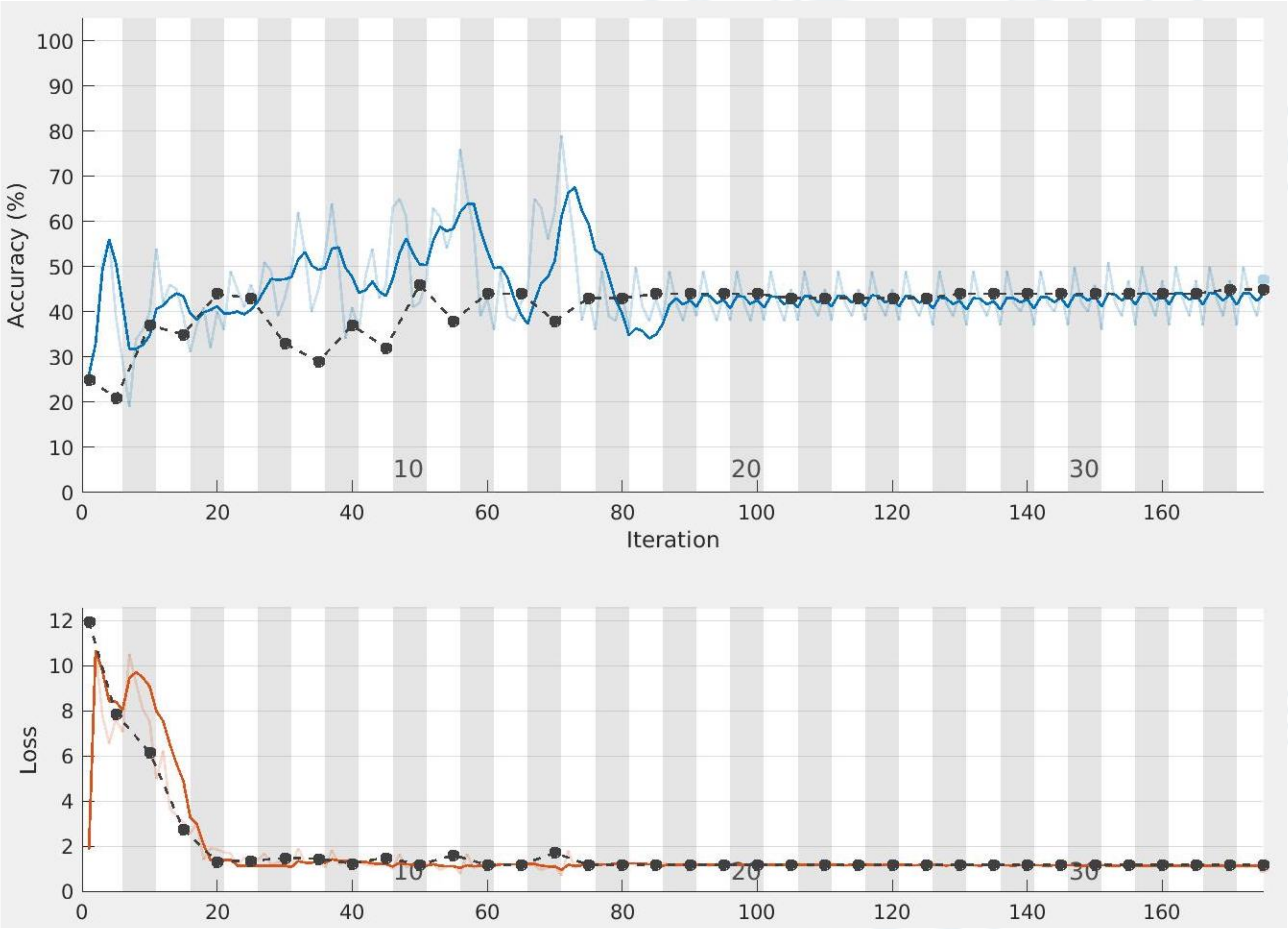
- Results for Dataset 3 (Grayscale Version)
- Results for Dataset 4 (Grayscale Version)
- Results of 50 Runs

Università degli studi di Trieste,
Department of Physics

Results for Dataset 3 (Grayscale Version)



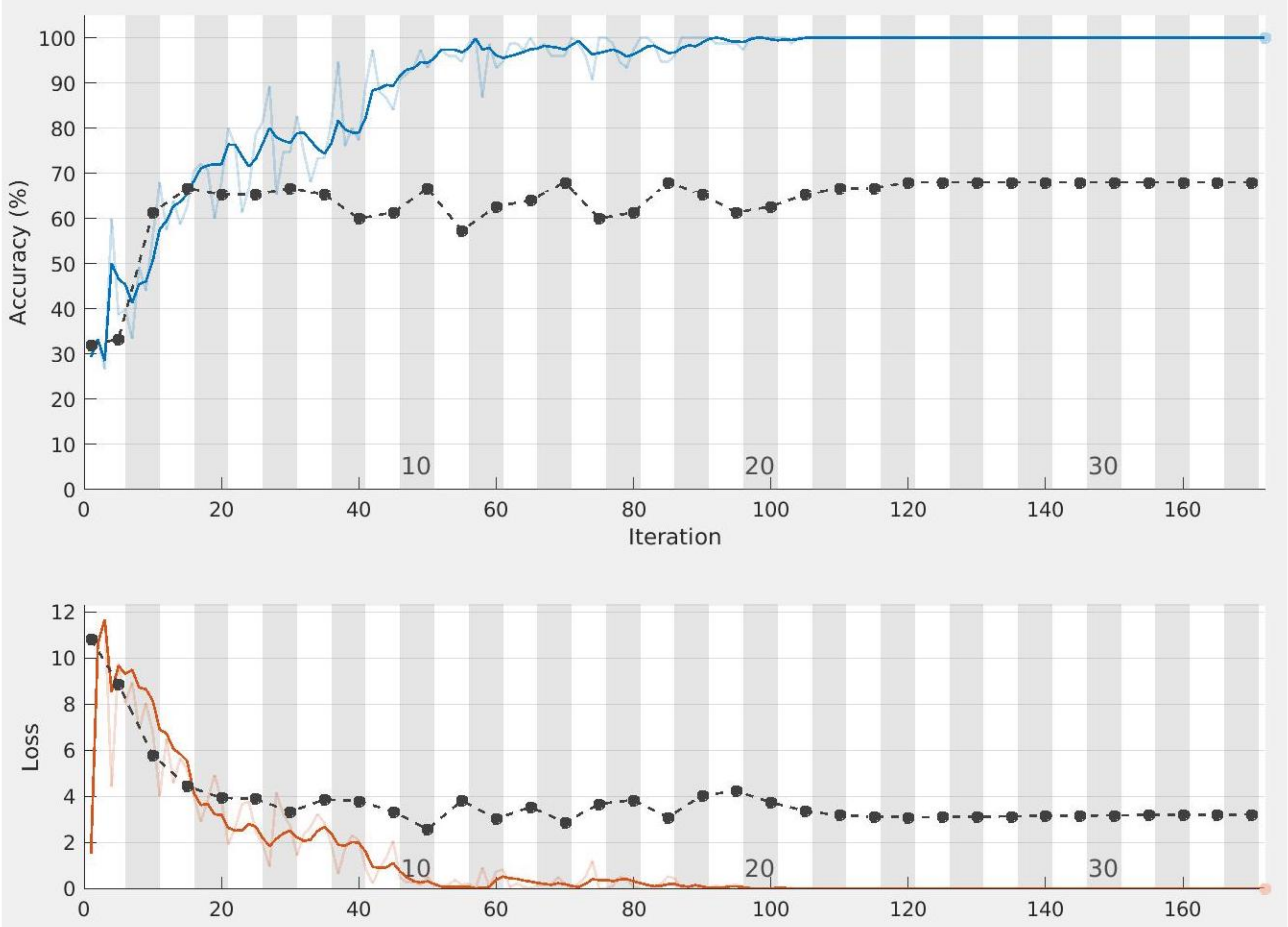
Four Classes - Accuracy: 52% External test: 38.67%



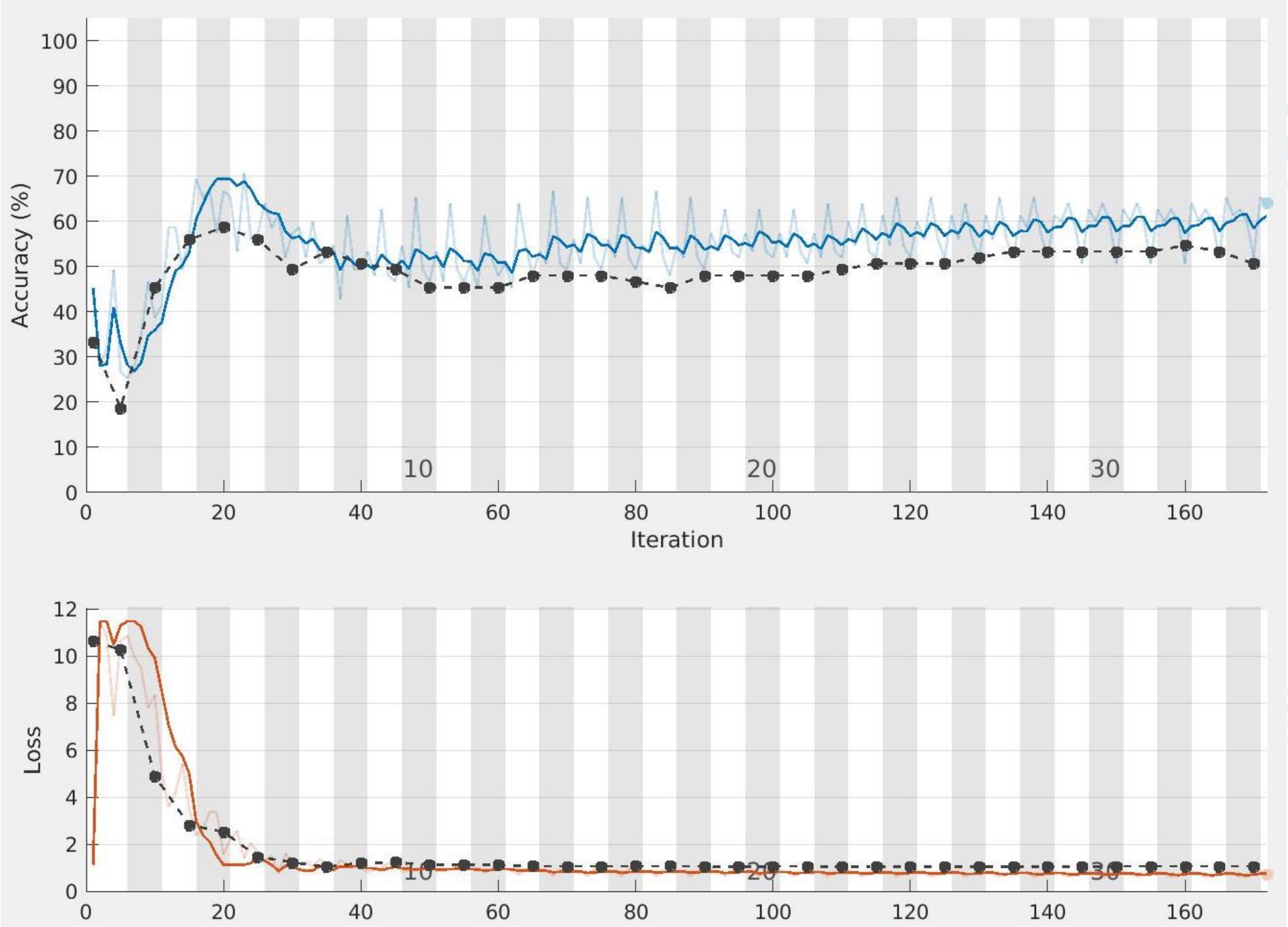
Four Classes - Accuracy: 43% External test: 52%

Università degli studi di Trieste,
Department of Physics

Results for Dataset 3 (Grayscale Version)

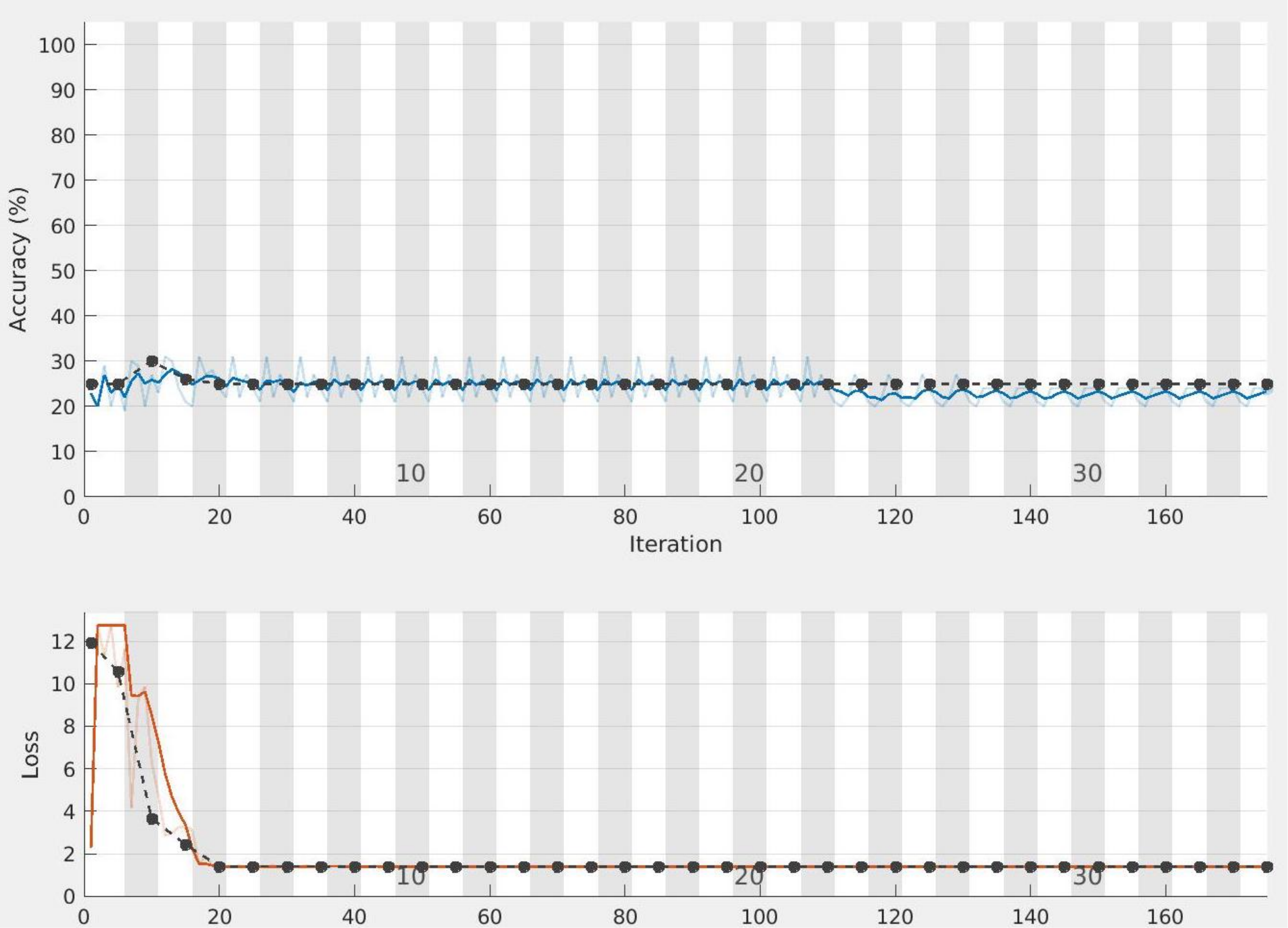


Three Classes - Accuracy: 68% External test: 56%

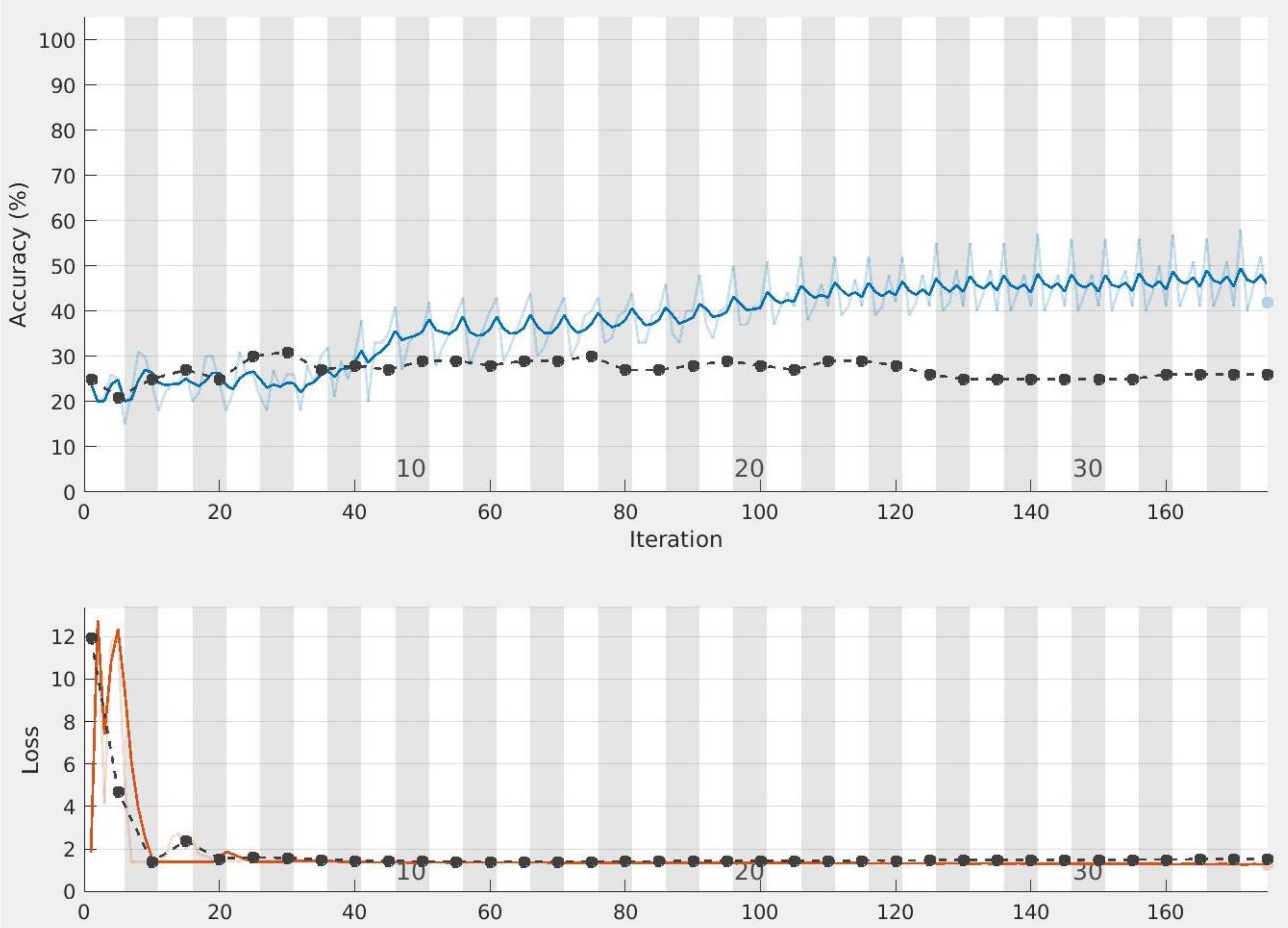


Three Classes - Accuracy: 53.33% External test: 66.67%

Results for Dataset 4 (Grayscale Version)



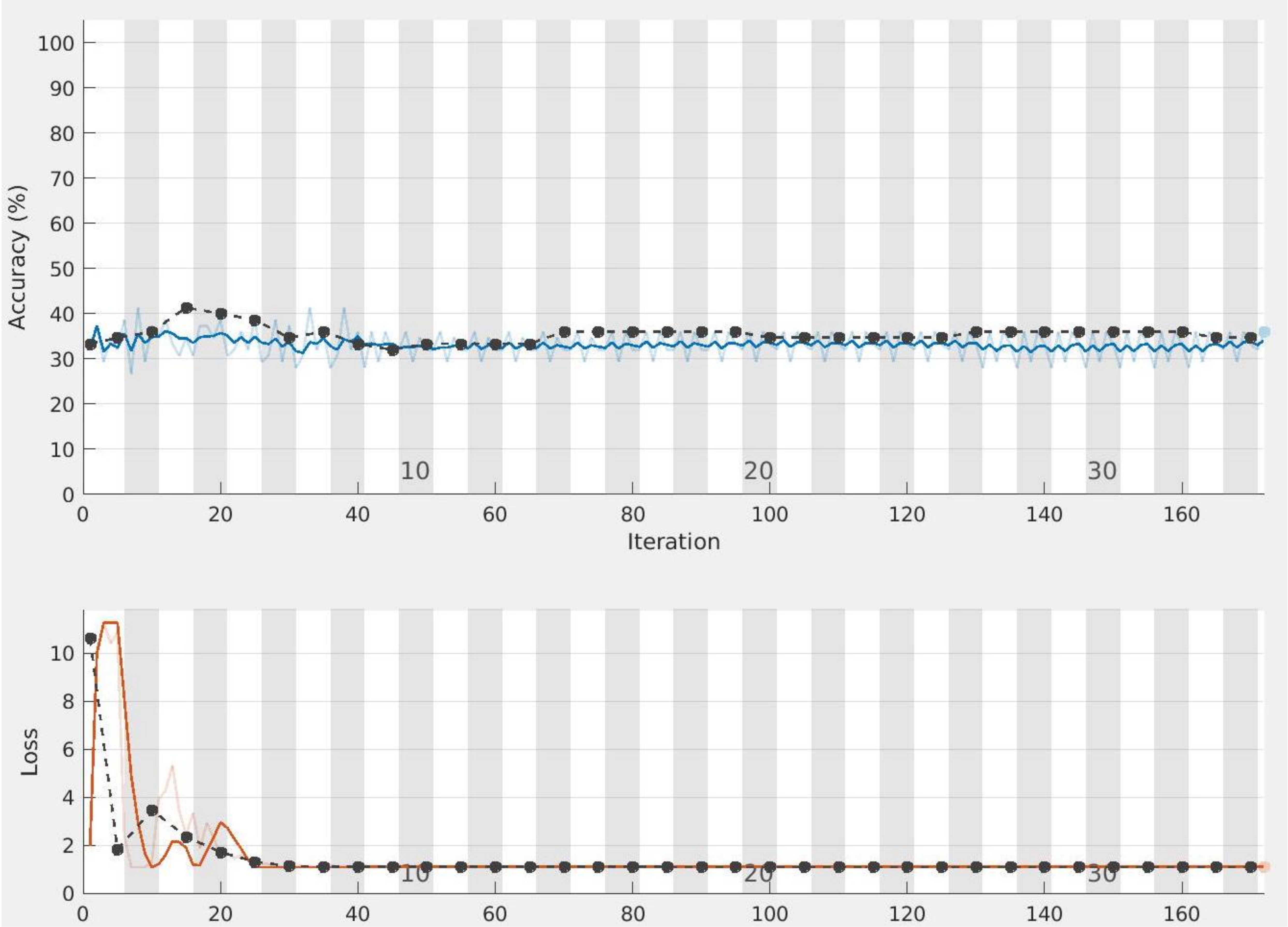
Four Classes - Accuracy: 25% External test: 33.33%



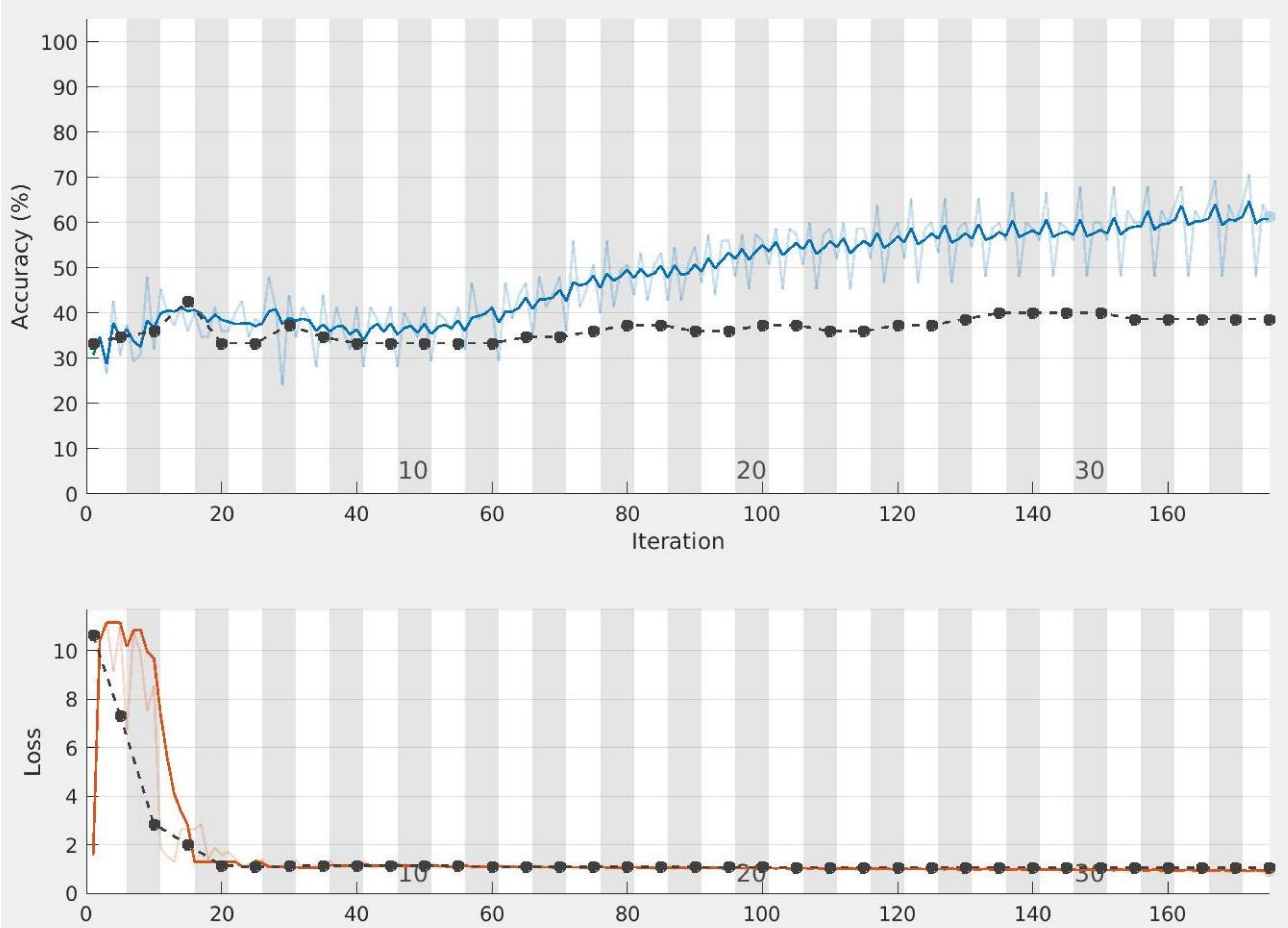
Four Classes - Accuracy: 28% External test: 40%

Università degli studi di Trieste,
Department of Physics

Results for Dataset 4 (Grayscale Version)



Three Classes - Accuracy: 36% External test: 33.33%



Three Classes - Accuracy: 40% External test: 30.67%

Università degli studi di Trieste,
Department of Physics

Results of 50 Runs

	4 Classes		3 Classes	
	Accuracy	External Test	Accuracy	External Test
Dataset 1 True Color	36.10	33.92	47.52	49.07
Dataset 1 Grayscale	36.10	35.78	48.29	47.25
Dataset 2 True Color	36.76	39.31	49.87	50.29
Dataset 2 Grayscale	36.40	37.73	49.81	49.52
Dataset 3 True Color	42.08	39.28	56.40	57.44
Dataset 3 Grayscale	43.68	42.64	56.69	55.95
Dataset 4 True Color	25.80	28.40	33.87	33.54
Dataset 4 Grayscale	26.54	26.80	33.84	33.89

Percentage variation of the outcomes of each training, referred to the random guessing probabilities: 25% and 33.33%.

	4 Classes		3 Classes	
	Accuracy	External Test	Accuracy	External Test
Dataset 1 True Color	44.40	35.68	42.57	47.22
Dataset 1 Grayscale	44.40	43.12	44.88	41.76
Dataset 2 True Color	47.04	57.24	49.62	50.89
Dataset 2 Grayscale	45.60	50.92	49.44	48.57
Dataset 3 True Color	68.32	57.12	69.22	72.34
Dataset 3 Grayscale	74.72	70.56	70.09	67.87
Dataset 4 True Color	3.20	13.60	1.62	0.63
Dataset 4 Grayscale	6.16	7.20	1.53	1.68

Università degli studi di Trieste,
Department of Physics

Conclusions and Perspectives

- A modified Convolutional Neural Network was successfully implemented in order to automatically classify the morphology of galaxies.

- An image selection process was set up and various pre-processing routines were tested.

- Using different datasets various results were obtained. These pointed out that High-Pass FFT Gaussian profile filtering provided the best results.

- These preliminary results obtained are very promising: Deep Learning techniques could be the key to the development of an automatic procedure for morphological classification of galaxies by extending the input dataset.

Thank you for your attention